УДК 371.048

МОДЕЛИРОВАНИЕ ЭКСПЕРТИЗЫ РАЗРАБАТЫВАЕМОГО ДИДАКТИЧЕСКОГО ТЕСТА

Н.С. Михайлова, М.Г. Минин

Томский политехнический университет
E-mail: mns@cts.tpu.ru

Дано описание организационно-педагогических условий экспертизы предтеста, способствующих разработке качественного инструмента для измерений в области учебных достижений. Рассмотрены объект, цели, этапы, возможные методы проведения адекватные целям. Подробно изложено содержание экспертизы с выделением критериев оценивания, предложены параметры оценивания.

1. Актуальность проблемы обеспечения качества дидактического теста

В последнее время все больше растет интерес к дидактическому тестированию как наиболее объективному методу оценки качества степени обученности. Но, как справедливо отмечают В. Шадриков и Н. Розина [1], «разработка тестов, проведение тестирования, обработка и представление результатов в стране пока не соответствуют международным стандартам в области тестирования».

Актуальность настоящих исследований вытекает из потребностей практики педагогического контроля по улучшению качества используемых тестовых материалов. Пожалуй, наиболее важным исходным моментом является определение существенных признаков той разработки, по результатам применения которой будут приниматься управленческие решения. Из множества известных стоит, с нашей точки зрения, придерживаться определения, данного А.Н. Майоровым [2]: «Тест – это инструмент, состоящий из квалиметрически выверенной системы тестовых заданий, стандартизированной процедуры и заранее спроектированной технологии обработки и анализа результатов, предназначенный для измерения качеств и свойств личности, изменение которых возможно в процессе систематического обучения». Если рассматривать тест как стандартизированный инструмент измерения качеств и свойств личности, то составными частями следует признать не только систему заданий, но технологию организации, обработки, анализа полученных результатов. В массовой педагогической практике нередко любой набор заданий, стандартизованных согласно требованиям тестовой формы, ассоциируется с тестом, исходя при этом из аналогии с привычными короткими контрольными заданиями. Такие псевдотесты сплошь и рядом публикуются в специальных сборниках, применяются в отечественной практике особенно часто внутри образовательных учреждений.

Исходя из понятия теста, экспертизе должны подвергаться не только задания, а вся измерительная система: и спецификация, и методическое оснащение теста, включающее, как минимум, набор требований, инструкций для участвующих в процедуре, сценарий тестирования, правила обработки и интерпретации результатов, а в случае использования компьютера – соответствие алгоритмов программного обеспечения тестологическим требованиям. Набор перечисленных материалов нельзя назвать тестом, так как не определена его окончательная структура, отсутствуют данные о системообразующих качествах, поэтому, уточняя понятие, будем придерживаться названия «предтест» [3].

В работах [2, 3] обращается внимание, прежде всего, на отсутствие массового отечественного опыта в разработке тестов и, как следствие, на необходимость достаточно подробных инструкций процедур экспертизы, «а если позволяют ресурсы, то и специальное обучение экспертов». Опыт подготовки тестологов-специалистов в области педагогических измерений в Томском политехническом университете показывает нерешенность как методических, так и организационных проблем в проведении экспертизы.

В предлагаемой нами педагогической модели сделана попытка по обобщению и совершенствованию процессов проведения экспертизы дидактических тестов. Первоначально была разработана имитационная модель экспертизы, при практической реализации отдельных этапов которой уточнены, скорректированы условия и процедуры экспертизы [4].

Следует отметить, что в работе мы не будем рассматривать математические модели поведения экспертов и математико-статистические методы анализа экспертных оценок, т.к. теория и практика экспертных оценок не относятся к области педагогики, хорошо исследованы в прикладной математике, социологии.

2. Основные этапы процедуры экспертизы предтеста

Выделены следующие стадии организации экспертного опроса:

1. формулировка цели экспертного опроса;
2. подбор руководителем основного состава рабочей группы;
3. разработка рабочей группой и утверждение у руководителя технического задания на проведение экспертного опроса;
4. разработка рабочей группой подробного сценария проведения сбора и анализа экспертных мнений (оценок);

5. подбор экспертов в соответствии с их компетентностью;
6. формирование экспертной комиссии (целесообразно заключение договоров с экспертами об условиях их работы и ее оплаты, утверждение состава экспертной комиссии);
7. проведение сбора экспертной информации;
8. анализ экспертной информации;
9. при наличии нескольких туров – повторение двух предыдущих этапов;
10. интерпретация полученных результатов и подготовка заключения для руководителя.

3. Основные методы экспертизы предтеста

Основная идея экспертных методов состоит в том, чтобы использовать интеллект людей, их способность искать и находить решение слабо формализованных задач. Однако особенность интеллектуальной деятельности людей состоит в том, что она во многом зависит от внешних и внутренних условий. Поэтому в технологии организации экспертных оценок специальное внимание уделяется созданию благоприятных условий и нейтрализации факторов, неблагоприятно влияющих на работу экспертов.

Существует масса методов получения экспертных оценок. Как показывает практика проведения экспертизы предтеста (централизованное тестирование), наиболее эффективен метод "Делфы" [5]. Если проводить оценку альтернатив в несколько туров, сообщая после каждого его полные итоги и сохраняя анонимность участников, то эксперты склонны не только критиковать, но и прислушиваться к критике, относящейся к ним лично. Обычно оказывается достаточно следующих четырех этапов:

1. раздача анкет, сбор оценок, их обобщенное представление с указанием разброса мнений;
2. сообщение итогов и запрос объяснений причин индивидуального отклонения от средней или медианной оценки первой итерации;
3. сообщение всех объяснений и запрос контраргументов на них;
4. сообщение возражений и запрос новых оценок альтернатив, если эксперт пожелает их изменить; нахождение окончательного итога.

Вся работа проводится под руководством отдельной управляющей группы, в которую входят системный аналитик и руководитель. Анонимность экспертов сохраняется до конца работы, а по желанию экспертов – и после ее окончания.

Поскольку взаимодействие между экспертами может, как стимулировать, так и подавлять их деятельность, в разных случаях можно рекомендовать методики экспертиз, имеющие различные степень и характер взаимного влияния экспертов друг на друга:

- анонимные и открытые опросы и анкетирование;
- совещания (комитеты, коллегии, комиссии, штабы);
- дискуссии (консилиумы, суды, ученые советы);
- деловые игры, мозговой штурм и т.д.

4. Подбор экспертов для оценивания

Структурообразующими частями модели экспертизы явились: цели, предмет, формы процедур, критерии оценки качества, интерпретация результатов. Но, прежде всего, необходимо было проанализировать и решить основную на сегодняшний момент проблему подбора экспертов. Мы исходили из того, что эксперт – это специалист в некоторой области, суждения и умозаключения которого в этой области обладают высокой степенью достоверности [5].

К специалистам-экспертам чаще необходимо предъявлять требования высокой практической подготовленности, способность к критическому анализу прошлого и настоящего и прогнозированию будущего, психологическая устойчивость, отсутствие склонности к соглашательству. Вместе с тем, от эксперта требуется эрудиция в смежных областях знаний: педагогике, психологии, теории педагогических измерений, предметной области знаний.

Количество необходимых экспертов зависит от целей создаваемого теста и квалификации экспертов. Руководит работой экспертной комиссии рабочая группа. Наиболее оптимальное количество экспертов в группе – 3–8 человек. В любом случае, эксперты назначаются, как минимум, руководителем организации, в которой будет использован тест.

Надо прямо сказать, что нет методов подбора экспертов, наверняка обеспечивающих успех экспертизы. Часто предлагают использовать методы взаимооценки и самооценки компетентности экспертов. С одной стороны, кто лучше может знать возможности эксперта, чем он сам? С другой стороны, при самооценке компетентности скорее оценивается степень самоуверенности эксперта, чем его реальная компетентность. Тем более, что само понятие «компетентность» строго не определено. Можно его уточнять, выделяя составляющие, но при этом усложняется предварительная часть деятельности экспертной комиссии.

Необходимо подчеркнуть, что подбор экспертов в конечном счете – функция рабочей группы.

В качестве независимых экспертов рассмотрено [2] три категории педагогов: администраторы, преподаватели, методисты по предмету. Если квалификация перечисленных категорий специалистов в области педагогических измерений недостаточна, обязательно включают тестологов. Наиболее успешно управленцы способны оценить четкость постановки задач тестирования, категорию испытуемых, соответствие программам обучения, организационные процедуры в сценарии испытаний. Преподавателям необходимо проанализировать полноту содержания, репрезентативность материала, выносимого на контроль, правильность формулировок заданий с точки зрения современных научных представлений, однозначность ответов.

5. Виды экспертиз предтеста

Эксперты должны руководствоваться общей целью своей деятельности, заключающейся в ответе на вопрос: «Какие изменения в рецензируемом материале будут способствовать корректному оцениванию содержания, уровня, структуры знаний у данных испытуемых?» Перед началом работы каждый эксперт должен ознакомиться с методикой и содержанием своей работы. Возможно, выполнение части экспертизы, исключая разделы, наименее отвечающие квалификации специалиста.

Такой подход позволяет использовать рекомендации В.С. Аванесова [6] по поэтапному проведению экспертизы исходного варианта теста, что повышает эффективность работы по его созданию. Ко всему прочему, в квалифицированном создании теста принимает участие несколько специалистов-разработчиков, отвечающих за осуществление разных этапов, результаты которых есть возможность планомерно корректировать по мере получения экспертных оценок. В случае суммирующей экспертизы (спецификации, системы заданий, методического оснащения) ошибки, заложенные на начальных этапах разработки теста, потребуют более серьезных изменений и повторного прохождения части процедур.

Исходя из вышеизложенного, мы рекомендуем следующий порядок этапов в разработке исходного варианта теста с параллельной поэтапной экспертизой по результатам каждого:

1. Определение цели тестирования, выбор вида теста по способу предъявления, обоснование подхода к его созданию.
2. Анализ и структурирование содержания учебной области, составление кодификатора, операционализация целей обучения.
3. Разработка спецификации теста, основой которой является тестовая матрица с характеристикой заданий, планирование системы заданий, выбор длины теста. Результаты работы по первым трем пунктам содержатся в разных разделах спецификации.
4. Создание заданий в тестовой форме, апробационное тестирование на небольшой выборке испытуемых для выявления ошибок, опечаток в тексте заданий, времени тестирования. Выбраковка заданий.
5. Создание методического оснащения.

6. Исходные материалы и документы экспертизы

Независимо от выбранного способа проведения оценивания качества теста экспертам должен быть выдан комплект материалов, в который входит:

- бланк о неразглашении содержания рецензируемых тестов;
- перечень пронумерованных требований стандартов и требований, на основе которых формировалось содержание тестового контроля автором;
- спецификация теста с набором инструкций, сценария, бланков;
- перечень учебных пособий, государственных образовательных стандартов (минимум содержания образования);
- тест (варианты);
- таблица правильных ответов с критериями оценки (в баллах) за выполнение каждого задания;
- бланки экспертизы с таблицами, пояснения по оцениванию, работе с бланками;
- договор;
- для компьютерного тестирования – структура и характеристика банка тестовых материалов, технология формирования и предъявления теста, демонстрационная версия банка, система оценивания результатов тестирования.

7. Разделы экспертизы предтеста

7.1. Качество спецификации

Основная часть спецификации – технологическая матрица, которая задает содержание проверки, уровни контролируемых учебных действий, соответствие стандарту и некоторые другие компоненты. Экспертиза спецификации необходима для обоснования оценки знаний по всему предмету на основе ответов по малому числу заданий теста. Экспертизу рекомендуется проводить по следующим направлениям с применением трехбалльной оценки (2,1,0).

- Обоснование выбора подходов к созданию теста, предлагаемых сфер применения теста (отвечает – частично отвечает – не отвечает целям тестирования).
- Соответствие разделов технологической матрицы разделам в государственном образовательном стандарте (ГОСе). При высокой степени несоответствия следует указать пропущенные разделы или те, которые не следовало выделять в самостоятельные блоки.
- Соответствие объема содержания разделов (программа, ГОС) количеству заданий на этот раздел в спецификации, т.е. пропорциональность числа заданий объему содержания разделов. При разработке возможно смещение акцентов в сторону тех разделов, по которым легче составить задания в тестовой форме. Для оценки правильности выбранных в спецификации пропорций подсчитывается процент, характеризующий отклонение мнения разработчика от мнения эксперта. Мера отклонения вычисляется в виде суммы разностей процентов по каждому разделу, каждая разность берется по абсолютной величине. Например, для 3-х разделов: $|30-20| + |26-55| + |44-25| = 10+29+19 = 58$ %. Среднее арифметическое равно 19,33 %.

Если полученный процент несоответствия менее 10, то точки зрения можно считать условно совпадающими. Несоответствие от 10 до 30 % означает, что точки зрения отличаются, но со спецификацией разработчика можно согласиться. Нес-

оответствие свыше 30 % требует привлечения дополнительных экспертов и углубленной работы над качеством содержания теста. Результаты экспертизы по данному пункту даются в баллах в соответствии с рекомендациями по таблице:

Таблица 1. *Оценка объема содержания разделов учебной дисциплины в спецификации*

Вид	Оценки		
	0 баллов	1 балл	2 балла
Пропорции	неправильные (свыше 30%)	удовлетворительные (от 10 до 30%)	правильные (менее 10%)

- Наличие ссылок на программы и другие нормативные документы.
- Распределение заданий по уровням усвоения знаний в соответствии с содержанием учебного предмета, требованиями к подготовленности испытуемых.

7.2. Анализ содержания отдельных заданий теста

Оценка качества отдельных индикаторов и результаты экспертизы содержания заданий оформляются в виде табл. 2.

Таблица 2. *Оценка качества отдельных заданий теста*

1.	№ задания	1	2	3	4
2.	№ правильного ответа				
3.	№ требования				
4.	Уровень базовости				
5.	Значимость содержания задания (2, 1, 0)				
6.	Максимальный балл за выполнение, варианты оценивания				
7.	Ожидаемый процент выполнения испытуемыми выборки				
8.	Ожидаемое время выполнения				
9.	Оценка дистракторов (2, 1, 0)				
10.	Неудачные задания (замечания)				
11.	Комментарии о качестве задания				
12.	*Тестологическая экспертиза:* Соблюдение грамматических правил, семантических требований (2, 1, 0)				
13.	Оформление (2, 1, 0)				
14.	Параллельность вариантов (2, 1, 0)				
15.	Система оценивания (2, 1, 0)				

По каждому пункту имеются инструкции по оцениванию. Например, для строки 13 «Оформление»: Оформление элементов тестового задания для быстрого поиска, технологически четкого выполнения задания (инструкция, номер задания, содержательная часть, варианты ответа, место для ответа): дизайн, требования эргономики, использование разных параметров шрифта (размер, выделение основных слов, использование курсива, цвета и т.п.). В этом же пункте следует отметить качество применения контекста: функциональность использования рисунков, диаграмм, аккуратность их выполнения, качество сделанных пояснений к рисункам, наличие лишней информации, подсказок.

7.3. Анализ теста как системы заданий

Содержание теста обычно значительно уже содержания учебной дисциплины. Чем полнее содержание, тем выше содержательная валидность, следовательно, и качественная оценка знаний с помощью данного теста. Оценка проводится по трехбалльной шкале по проценту объема содержания контроля в тесте от общего объема содержания предмета.

Таблица 3. *Оценка объема содержания учебной дисциплины в тесте*

Оценки		
0 баллов	1 балл	2 балла
не полное менее 40 %	полное 40...70 %	очень полное свыше 70 %

- Соответствие планируемой пропорциональности содержания в спецификации реальному содержанию теста.
- Так же оцениваются пропорции распределения по деятельностному аспекту – уровням знаний.
- Полнота охвата требований к уровню подготовки испытуемых. В основе получения экспертной оценки лежат результаты, помещенные в третьей строке табл. 2.
- Соответствие содержания заданий теста перечню требований к уровням подготовки, планируемым в спецификации теста. Степень несоответствия теста задачам тестирования определяется подсчетом процента заданий, не предусмотренных в спецификации теста. Например, знание классов неорганических веществ в спецификации планировалось проверить с помощью одного задания, а в тесте их получилось 3. Умение составлять формулы веществ по названию оценивалось двумя заданиями, а в тесте только одно. Сопоставление планируемых и реальных количеств заданий в тесте проводится путем вычисления разностей. Таким образом, подсчитывается общее число заданий, несоответствующих спецификации теста. Для данных примеров это число **a** будет:

  $$\mathbf{a}=(3-1)+(2-1)+...$$

  Число **a** делится на общее число заданий теста и умножается на 100 %. Оценка соответствия дана в табл. 4.
- Соответствие содержания теста содержанию образовательных стандартов по предмету.

Таблица 4. *Оценка структуры теста*

Вид	Оценки		
	0 баллов	1 балл	2 балла
Соответствие задачам	неудовлетворительное (более 30 %)	удовлетворительные (10...30 %)	полное (менее 10 %)

7.4. Оценка методического оснащения теста

Экспертам необходимо оценить представленное методическое оснащения, учитывая уровень использования теста:

- Степень соответствия (адекватность) инструкций форме и содержанию заданий.
- Полноту информации, заносимую при регистрации тестируемым в соответствии с целями использования. Какие данные упущены, избыточны?
- Насколько удобна испытуемому предлагаемая форма занесения ответа. Что можно предложить для устранения непродуктивных затрат и случайных ошибок при оформлении регистрационных данных, проставлении ответа?
- Дизайнерское оформление бланков, использование шрифта, цвета, наличие и размещение примеров заполнения бланков, инструкций для испытуемых.
- Примерное время, которое потребуется испытуемому для вдумчивого чтения инструкции. Предусмотрено ли объяснение инструкций ведущим? Целесообразно ли выполнение тренировочных заданий или размещение их на бланках?
- Создание мотивации деятельности тестируемых.
- Уровень детализации деятельности участвующих в тестировании. Соблюдена ли точность в пошаговой деятельности ведущего? Какие действия ведущего, испытуемого не оговорены (что могут, что обязаны, что запрещено делать)?
- Есть ли указания по внесению исправлений? Насколько понятно они составлены?
- Точность в списке необходимого оборудования, порядке размещения и предъявления инструкций, бланков, приборов.

7.5. Оценка алгоритмов программных средств компьютерной поддержки.

На настоящем этапе практически все создаваемые тестовые системы имеют компьютерную поддержку, которая может использоваться как для создания теста из банка тестовых заданий, для тиражирования уже созданных тестов (бланковое, компьютерное тестирование), проверки ответов, получения и интерпретации результатов, статистической обработки полученных результатов и т.д. Как показывает практика, использование компьютерных технологий в тестировании редко опирается на научно-обоснованные алгоритмы и процедуры, что приводит к низкому качеству и эффективности результатов измерения, несмотря на валидность составляющих тест заданий.

7.6. Обобщающие выводы и рекомендации

Эксперт приводит свое общее впечатление о содержании теста. Здесь высказываются все сомнения и пожелания эксперта, его рекомендации по улучшению качества теста. Возможна оценка соотношения заданий, проверяющих знание теории предмета и его практики. Желательно выявить задания, предназначенные для проверки системы понятий, а также задания интегрального характера, предназначенные для проверки умений учащихся обобщать знания по различным разделам предмета. Возможно ли их включение в предлагаемый тест в соответствии с целями контроля?

Завершается общим заключением эксперта о возможности использования рецензируемого теста для предложенных в спецификации целей и испытуемых.

8. Общие выводы

Предложенная модель служит основой для организации отдельных этапов экспертизы разрабатываемых контрольно-измерительных материалов в центре тестирования ТПУ. На ее основе созданы методические указания по проведению экспертизы отдельных тестовых заданий, разработана организационная схема экспертизы. Модель послужила основой выполнения ряда научно-педагогических исследований на кафедре педагогических измерений ТПУ, при выполнении ВКР обучающимися по программе дополнительного образования «Тестолог». Следующим шагом в реализации модели является подготовка экспертов на кафедре педагогических измерений ТПУ в рамках курсов повышения квалификации с учетом современных условий.

Несомненно, точные критерии в методике экспертизы могут рассматриваться как опосредованные требования к качеству теста, что способствует повышению эффективности конструирования и применения дидактических измерителей. Кроме администраторов педагогических учреждений, экспертов, содержание методики заинтересует любого педагога, нуждающегося в выборе или разработке качественного инструмента контроля, в доработке «сырых» комплектов заданий до приемлемого уровня качества.

Большая часть методики, разработанной на модели экспертизы, может быть использована для определения содержательной валидности теста, т.к. пригодность теста определяется, в основном, путем качественного оценивания с привлечением экспертов.

СПИСОК ЛИТЕРАТУРЫ

1. Шадриков В., Розина Н. Централизованное тестирование: проблемы и перспективы // Высшее образование в России. – 2000. – № 1. – С. 27–31.
2. Майоров А.Н. Теория и практика создания тестов для системы образования. – М.: Интеллект-центр, 2001. – 296 с.
3. Челышкова М.Б. Теория и практика конструирования педагогических тестов. – М.: Логос, 2002. – 432 с.
4. Михайлова Н.С. Организация и проведение экспертизы системы тестовых заданий (предтеста) // Качество образования: менеджмент, достижения, проблемы: Матер. VI Междунар. научно-метод. конф. – Новосибирск: Изд-во НГТУ, 2005. – С. 331–333.
5. Соколов В.Н. Педагогическая эвристика: Введение в теорию и методику эвристической деятельности. – М.: Аспект Пресс, 1995. – 225 с.
6. Аванесов В.С. Композиция тестовых заданий. Учебная книга. 3-е изд., доп. – М.: Центр тестирования, 2002. – 240 с.